Supplemental Information

Measuring rhythms of vocal interactions: A proof of principle in harbour seal pups

Marianna Anichini, Koen de Reus, Taylor A. Hersh, Daria Valente, Anna Salazar-Casals, Caroline Berry, Peter E. Keller and Andrea Ravignani.

Correspondence: Marianna Anichini
e-mail: marianna.anichini@mpi.nl

Method S1. Animal model and housing conditions

Nine orphaned, emaciated, and/or weakened harbour seal pups *(Phoca vitulina*) comprise our study group. These wild-born seal pups were admitted to rehabilitation at Sealcentre Pieterburen (Pieterburen, the Netherlands) before being released back into their natural environment. The animals originated from different Wadden Sea islands (e.g., Schiermonnikoog, Ameland, Vlieland) and from the Dutch mainland coast. The study took place in the intensive care units of the Sealcentre, where the animals were housed in enclosures consisting of a swimming pool and a plateau on which the animals could rest (Figure S1). One seal (pup I) was housed alone and exposed to a playback experiment during the summer of 2017 (1-way interaction). The remaining eight pups (A–H) were housed in pairs and vocalised spontaneously (i.e., in the absence of any playback stimulus) during the summer of 2020 (2-way interaction). Prior to each recording, we marked one of the two pups in the pair with a non-toxic coloured marker to distinguish the individuals later on in the analysis (Figure S1A). The intensive care units received natural light.

Method S2. Data collection and experimental procedure

All audio recordings were collected without human presence in the intensive care units. The lone seal (pup I) was recorded in two behavioural contexts: alone and in interaction with a recorded/playback partner (1-way interaction). Both context datasets were the focus of past works [1,2] and have been re-analysed in the present study. This pup was estimated to be seven days old upon admission to the Sealcentre. During the alone context, the pup's vocalisations were recorded between 9 and 27 days of age for twenty sessions of ten minutes each at a distance of 0.5–2 m from the pup (Table S1).

Between 29 and 37 days of age, the pup was recorded for five days and exposed to five playback sessions (Table S1). Each playback session lasted approximately 15 minutes and consisted of 18 trials (concatenated in random order) separated by 12 seconds of silence. Each trial was a unique combination of three experimental variables, including tempo (corresponded to the average IOI of a sequence: slow = 2418 ms, medium = 1983 ms, and fast = 1548 ms), rhythmicity (isochrony, random), and playback caller identity (self, a pup from the same Wadden Sea region, a pup from a different Wadden Sea region). Each trial consisted of 21 identical vocalisations. Playbacks were broadcasted from a JBL Flip 2 Bluetooth speaker (frequency response: 100 Hz–20 kHz; JBL, Los Angeles, CA), hidden from the seal's sight, and connected to an iPhone 5S (Apple, Cupertino, CA). The calls of the pup alone, the playback calls, and the vocal responses of the pup were recorded as .wav files (48 kHz, 24 bits) using a unidirectional microphone Sennheiser ME-66 (frequency response: 40 Hz–20 kHz; Sennheiser electronic GmbH & Co. KG, Wedemark, Germany), connected to a Zoom H6 digital recorder (Zoom Corporation, Tokyo, Japan).

The remaining eight pups (A–H), housed in four pairs, were recorded during the first two weeks of July 2020 in two behavioural contexts: when their partner was silent and when their partner was vocalising (i.e., the 2-way interaction). Recording sessions lasted approximately 3–4 hours and were conducted over a variable number of consecutive days per pair (Table S1). The variation in the number of recording sessions between pairs of pups was due to the length of time the pups were housed in the same room (subjected to veterinary decisions). Three tripods, each holding two semi-directional microphones (Sennheiser ME 64, frequency response: 40 Hz–20 kHz; Sennheiser electronic GmbH & Co. KG, Wedemark, Germany), were placed at three corners of the pool (Figure S1B). After calibration, the microphones were fixed at a height of 150 cm and connected to a multitrack recording device (Zoom F8; Zoom Corporation, Tokyo, Japan), and the recordings were collected as .wav files (48 kHz, 16 bits). A video camera (Canon LEGRIA HF G30 HD, 1080p resolution, 50 frames per second; Canon Inc., Tokyo, Japan), positioned next to one of the tripods, was used to record the video sequences in MP4 format. The video camera was also connected to the multitrack recording device to facilitate audio/video synchronisation.

Method S3. Data pre-processing and acoustic metric extraction

For the alone and 1-way interaction contexts, audio recordings were manually annotated in the acoustic software Praat 6.1.42 (http://www.praat.org/, [3]) and the annotations were imported into Python using the *TextGridTools* package [4]. For the silent partner and 2-way interaction contexts, the audio data were synchronised with the video data, allowing the identity of the calling pup to be determined for each interaction. A total of 373 vocalisations were attributed to a specific pup in the 2-way interactions. Prior to analysis, the audio files were resampled (from 48 kHz to 8 kHz) and filtered (with a high-frequency bandpass filter at 0.1 kHz) using SOX Sound eXchange (http://sox.sourceforge.net). This pre-processing procedure had no impact on the quality of the pups' calls which were subsequently manually annotated, creating labels at the start and end of each call, using Audacity software (v.2.3.3, [5]). All video recordings were annotated using BORIS software (v.7.9.19, [6]) in order to attribute the calls to the correct individual. A custom-written R script allowed us to extract the call parameters, onsets, and offsets (Figure 1) for each vocalisation and, from these, we calculated the rhythmic metrics used in subsequent analyses.

Method S4. Circular statistics

The circular statistics analyses were performed using R statistical software (v.3.6.3, [7]) and RStudio (v.2021.09.1, [8]), using the R package *circular* [9]. Following [2], we considered the values of the response phases as circular data falling between 0° and 360°. From the descriptive statistics, we obtained the circular mean (μ), median, and standard deviation (*v*) for all four behavioural contexts (Table S3). In addition, we calculated the 'mean resultant length' (ρ), which is a measure of the spread around a circle of the response phase data and ranges from 0 (large spread) to 1 (small spread) [10,11]. Using circular plots, we visualised the distribution of the response phases and corresponding circular means for all behavioural contexts (Figure 2, Figure S3). Afterwards, we tested for uniformity against a specified direction for the unimodal peak with a V-test or, in other words, tested the null hypothesis of a unimodal departure against two specified, alternative μ values: 0° (to test for synchrony) and 90° (to test for asynchrony). We specifically chose to test 90° following results from past work [2]. Following Landler *et al.* [12], we did not compute the V-test with the exact value of the circular mean value (thus avoiding type I error), and used the confidence intervals (CI) of μ to interpret the results: a narrow CI that includes the tested value 90° would indicate support for our alternative hypothesis of asynchrony. We computed the CI following the function (ConfIntLS) given by Pewsey *et al.* [11]. Since the Rayleigh and the V-test are theoretically based on the von Mises distribution (i.e., assuming a von Mises deviation from uniformity [12]), we tested if the response phases in all four behavioural contexts followed a von Mises distribution using one-sample Watson tests (Table S5). Like a normal distribution, a von Mises distribution is symmetrical and unimodal with a concentration around the mean falling away to give a bell shape [12]. With deviations from uniformity (null hypothesis von Mises distribution rejected), we used Kuiper's test, Watson's test, and Rao's spacing test, to confirm the p-value obtained from the Rayleigh test (Table S6) as suggested by Landler *et al.* [12].

Method S5. Categorical rhythm analysis

In the alone and silent partner contexts, IOIs were calculated for each bout as the difference in adjacent call onsets made by the vocalising pup (Figure 1A/B). For the 1-way interaction context, we calculated IOIs in two ways: 1) as the difference in the pup's adjacent call onsets disregarding the playback stimuli, and 2) as the difference between all adjacent call onsets (playback or pup) within interactions (Figure 1C). We used the latter approach to calculate IOIs in the 2-way interaction context as well (Figure 1D).

The null ratio distributions were generated by randomly sampling two IOIs from the empirically observed range for each pup/condition (Table S8) and calculating the IOI ratio. This process was repeated 100,000 times. From the simulated null ratio distribution, we created 1,000 resampled (without replacement) distributions that had the same number of ratios as the empirical distribution. We then compared the 1,000 resampled distributions with the empirical distribution using one-sample Kolmogorov-Smirnov (KS) tests and recorded the final range of the D statistic (across all 1,000 comparisons) and the average p-value (Table S9). For pups with at least 10 ratios in multiple contexts, we also used two-sample KS tests to determine whether the pup-specific ratio distributions differed in different behavioural contexts.

For the categorical rhythm analyses, the on-integer ratio ranges were centred around seven small integers: 1:4 (0.2), 1:3 (0.25), 1:2 (0.333), 1:1 (0.5), 2:1 (0.667), 3:1 (0.75), and 4:1 (0.8). The corresponding off-integer ratio ranges, as well as the bounds for each range, are described in Table S7. This resulted in 28 "bins" (denoted by the white and dashed black lines in all ratio distribution plots), and we counted the number of ratios that fell into each bin for each pup/context. Counts were normalised by bin size (i.e., the size of each bin's range on the x-axis), and on- and off-integer counts were then compared for each small integer ratio using paired Wilcoxon signed-rank tests. All analyses were conducted using R statistical software (v.3.6.1, [7]) and significance was generally set at $p<0.05$. The sole exception was for pairwise comparisons of pup I's empirical ratio distribution across three contexts (alone, 1-way interaction disregarding playback, 1-way interaction when pup I responds), in which case the p-value was Bonferonni corrected to $p < 0.017$ (Table S10).

Method S6. Time series analysis: Granger causality

The causality between two variables X and Y can be tested by measuring the ability to forecast the values of a time series using the values of a second time series [13]. Granger's postulate on causality is based on two assumptions: effects happen after causes, and a cause gives information about future values of the effect [13]. Time series X 'Granger-causes' time series Y if predictions of Y values based on its own past values and on X past values are better than predictions of Y based only on its own past values. To assess whether the timing of a calling pup can be predicted by an external stimulus (either a second calling pup or a playback call), we computed the Granger Causality Test using the R package *lmtest* [14]. In all cases, the test was conducted using the onset timing of the calls and lag measures from one to five. We chose this range of lag values for two reasons. First, selecting high lag values implies a loss of observations [14]; therefore, the lower the number of observations, the smaller the lag value that should be used. Secondly, given that most of our samples consisted of series of five to ten paired calls onsets, using the same lag values in all the analyses ensures comparability between the different datasets we considered. The test was run either in one direction, in cases where pup I was calling after a playback stimulus, or bidirectionally, for each pair of pups calling together. Before running the causality test, we used the augmented Dickey Fuller test for unit roots [15] to verify whether the time series involved in the study were stationary and we performed a first-order differencing when the stationarity assumption was not met. Bar plots were generated using the R package *ggpubr* 0.4.0 [16].

Method S7. Time series analysis: ADAM

The ADaptation and Anticipation Model (ADAM; Figure S2) was used to estimate seal pups' use of reactive error correction and predictive processes in a 2-way interactive scenario between pups A and B and 1-way interaction of pup I with recordings in five playback sessions. The model parameters of interest were phase correction ('alpha') and period correction ('beta') from ADAM's adaptation module, temporal prediction/tracking ('delta') from the anticipation module, and anticipatory error correction ('gamma') from the joint module. Parameters were also estimated for two sources of noise, one that represents variability in a 'timekeeper' in the central nervous system and the other representing 'motor' noise in the peripheral nervous system. We do not report results for the timekeeper and motor noise estimates (but nevertheless included them to account for associated variance in the data).

ADAM parameters were estimated for each pup using MATLAB scripts adapted from [17]. The estimation procedure used the bounded Generalised Least Squares method [18, 19], which harnesses the analytic efficiency of matrix algebra applied to behavioural time series data [20].

A series of steps were taken to screen and select seal pup vocalisation data for use with ADAM. The objective of these steps was to identify pup calling bouts that were long enough to yield reliable parameter estimates. Previous work suggests that time series with less than 20 event pairs are insufficient for reliable parameter estimation with the statistical techniques implemented in ADAM [20].

For the session featuring multiple pups, bouts were grouped by pup pairs. Pups in one pair (pups A & B) produced a relatively large number of interactive bouts, while another pair (pups E & F) was intermediate, and two pairs (pups C/D and pups G/H) displayed less vocal interaction. ADAM parameter estimation was therefore restricted to pups A and B for the 2-way interaction context sessions. All five playback sessions featuring pup I were analysed.

Interactive bouts were shorter than 20 events for all data sets, and we, therefore, concatenated these short bouts to produce longer sequences suitable for parameter estimation. Time intervals between the final call of one bout and the first call of the next bout varied considerably. To reduce the effect of this variation, inter-bout intervals longer than 60 seconds were filtered out (lowering the critical value would result in sequences that are too short for reliable parameter estimation). A total of 15 inter-bout intervals were removed from the full data set. This resulted in longer, final sequences ranging from 21 to 91 events (see Table S12).

The effects of sequence length and the concatenation procedure on the reliability of parameter estimates were examined by simulating synchronisation with short and long sequences with the same combinations of parameter settings, and then testing how well these settings could be recovered by the parameter estimation procedure applied to these different simulated sequences (see Discussion S1). The results of this test indicated that the reliability of ADAM parameter estimates for present purposes is not seriously undermined by differing sequence lengths or by concatenation.

Once the data of interest were extracted, they were transformed into a format appropriate for entry into ADAM. This format consists of three time series of equal length. The modelling procedure treats one series as reference time intervals associated with a pacing sequence ('inter-stimulus intervals'), another series as time intervals associated with movements produced by an individual synchronising with the pacing sequence ('inter-response intervals'), and a final series consisting of the 'asynchronies' between the onsets of corresponding events in the pacing sequence and the response sequence (stimulus onset times are subtracted from response onset times, yielding positive asynchronies if responses occur after stimuli).

Parameter estimates were obtained using the adaptation-only and full ‘joint’ versions of ADAM. Anticipatory error correction (gamma) estimates are restricted to the range 0–1, as parameter range restriction generally allows for more reliable estimates [20,21] and values outside this range for this parameter, in particular, would complicate the interpretation of results. These estimates reached the upper bound of 1 for both the original and permuted data, meaning that results cannot be interpreted conclusively due to the equivalence of estimates and are thus not shown in Figure 4 in the main article. The quality of the fits of the adaptation-only model and the joint model to the data was evaluated by comparing log-likelihood estimates (LLE) for the two models across datasets in a paired-samples t-test. Fits for the adaptation model (mean LLE = -271, SD = 141) and joint model (mean LLE = -255, SD = 135) did not differ significantly ($t(6)= -0.611$, $p = 0.564$).

To obtain estimates for two agents within a pair, the estimation procedure is run twice: once with one agent serving as the focal individual and the other as the external reference, and then with these roles reversed. In the present case, this reciprocal estimation procedure was applied only to data from interacting pups. Similar parameter estimates for each pup within a pair indicate a symmetrical pattern of mutual influence, whereas different estimates for both pups indicate asymmetrical influence, which could indicate leader-follower roles or processes enhancing the temporal distinction of calls. Note that the relative magnitude of parameter estimates is more meaningful than the absolute magnitude for the present, concatenated data (because larger absolute values could indicate either a large change in rate or skipped events, and it is difficult to distinguish between these two possibilities).

To test the reliability of the obtained ADAM parameter estimates, they were compared against corresponding values that were estimated for randomly permuted data. For this comparison, we ran 10,000 permutations in which the order of events within a sequence was randomly shuffled. Specifically, the rows of the data matrix (with each row containing a reference inter-stimulus interval, an inter-response interval, and an asynchrony) were randomly ordered 10,000 times for each sequence before parameter estimation procedures were applied for the two versions of ADAM. Each obtained parameter estimate for the real (unpermuted) data was then converted to a z score based on the mean and standard deviation of the corresponding 10,000 permuted data estimates, and the p-value for the z score was calculated. It was assumed that obtained parameter estimates for which 2-tailed $p < 0.05$ were unlikely to have arisen by chance and therefore were interpretable. Given the exploratory nature of the modelling exercise, we also discuss estimates that are significant at the 1-tailed level.

Discussion S1. Effects of concatenating bouts on ADAM parameter estimates

Seal pups produced sequences of interactive calls that were deemed too short for each bout to be modelled individually (for guidelines, see [20]). We therefore concatenated these interactive bouts to produce longer sequences. The current simulations assess the potential effects of this concatenation process on the estimates of the ADAM parameters. The concatenation procedure might, for example, introduce a bias in the estimates due to discontinuities at the transition points between bouts (e.g., jumps in tempo or timing adjustments that are not consistent with the combination of the parameter settings of the preceding events).

In the present tests, we simulated the task of one agent synchronising with an external pacing sequence (analogous to the playback session in 1-way interaction). This single agent case can be assumed to be equivalent to two interactive agents to the extent that the ADAM modelling procedure consists of assigning one agent as the referent (i.e., the external pacing sequence), and then repeating the process with the other agent as the referent, thus obtaining separate parameter estimates for each of the two agents.

The simulations addressed two timing regimes, one in which tempo remains steady and the other in which tempo changes gradually. To match the wide range of time intervals present in the seal pup data, tempo varied within a range spanning intervals of 1 to 1000 ms.

Two sets of simulations were performed. Set 1 compared parameter estimates for long sequences versus averaged estimates for short sequences (i.e., the segmented long sequences). Set 1a tested multiple long sequences for each tempo regimen, while set 1b used a single sequence for each regimen, as this was more convenient for comparison with set 2. Set 2 simulates synchronisation with long sequences and short sequences (again, segments of the long sequences) and then concatenates the simulated short time series of IOIs and asynchronies prior to parameter estimation. Set 1 therefore addresses the effects of sequence length on parameter estimates, while set 2 addresses sequence length and potential discontinuities that occur due to the concatenation procedure. Effects of concatenation (independent of sequence length) can be gauged by comparing the two sets of simulations.

Simulation set 1: Effects of sequence length

Set 1a used multiple sequences for each tempo regimen. Pacing sequences were generated by first specifying the number of IOIs (64 or 96) and then assigning each interval a value that, for steady tempo sequences, was the mean of the overall range (500.5 ms) or, for tempo-changing sequences, got progressively slower and faster in alternation in accordance with groups of prespecified length (8+8+8+8+8+8+8+8; 6+10+6+10+6+10+6+10; 12+12+12+12+12+12+12+12; 10+14+10+14+10+14+10+14). This resulted in two sequences of 64 intervals and two sequences of 96 intervals for each tempo regime. These sequences were then segmented into shorter 'child' sequences based on pairs of groups from the original longer 'parent' sequences (8+8; 8+8; 8+8; 8+8; 6+10; 6+10; 6+10… etc). Set 1b was like set 1a but used just a single, longer parent sequence with 128 IOIs with a repeating value of 500.5 ms for the steady tempo and mixed grouping for the sequence with tempo changes (8+8+8+8+8+8+8+8+6+6+10+10+6+6+10+10).

Synchronisation with the long parent sequences and short child sequences in each of the two tempo regimens was then simulated for all possible combinations of ADAM parameter settings (within pre-specified ranges). The adaptation-only version of ADAM (with parameters 'alpha' and 'beta') was used for steady tempo simulations, and the full 'joint' model (with parameters 'beta' 'delta' and 'gamma') was used for tempo-changing simulations. Parameter settings ranged from 0 to 1 in 0.1 steps. Timekeeper noise was set to 10 and motor noise was set to 0 in all simulations. One hundred runs were executed for each parameter combination for long parent and short child sequences in each tempo regimen.

Following each simulation with a given combination of parameter settings, a procedure was executed to estimate the parameter values from the simulated time series of IOIs and asynchronies. The resulting parameter estimates were then averaged across all sequences and run for each combination of original parameter settings separately for the long parent and short child sequences.

The relationship between long and short sequences was examined for two behavioural outcome measures: the mean asynchrony and the variability (standard deviation) of asynchronies, as well as the ADAM parameter estimates for the two tempo regimens. Scatterplots of the relationships for each measure and estimate are shown in Figure S7 (simulation set 1a) and Figure S8 (set 1b). Canonical correlation analyses were performed to quantify the strength of the relationships across parameters using the MATLAB function *canoncorr*. A canonical correlation analysis models the associations between two multidimensional datasets, each containing several different variables (here, parameter estimates), by finding linear combinations of corresponding variables in each dataset that have the highest correlation (and are statistically independent of other pairs of canonical variables in the dataset). These analyses revealed strong correlations between parameter estimates for long and short sequences for both simulation sets for the adaptation model (set 1a: Wilks' lambda = 1.99e-07; $F(16,345.86) = 3354.70$; $p < 0.001$; set 1b: Wilks' lambda = 2.34e-05; $F(16,345.86) = 687.18$; $p < 0.001$) and the joint model (set 1a: Wilks' lambda = 2.87e-05; $F(16, 44705.19) = 82{,}881.98$; $p < 0.001$; set 1b: Wilks' lambda = 2.59e-04; $F(16, 44705.19) = 38{,}920.15$; $p < 0.001$). Based on these results, we conclude that the reliability of ADAM parameter estimates is not undermined by differing sequence lengths.

Simulation set 2: Effects of sequence length and concatenation

Set 2 addresses the potential effects of sequence length and potential discontinuities due to concatenation by first simulating synchronisation with long sequence and short sequences and then concatenating the simulated short time series of IOIs and asynchronies prior to parameter estimation. Set 2a uses four sequences and set 2b uses one longer sequence. Scatterplots of the relationships for each dependent measure and parameter estimate are shown in Figure S9 (simulation set 2a) and Figure S10 (set 2b).

A conspicuous difference from set 1 is the lowering of the estimated parameter values for short concatenated (part) sequences (compare scales on vertical axes). This suggests that discontinuities led to a reduction bias whereby the parameter values were underestimated. Apart from this consistent bias, which affects the absolute parameter values, robust relationships emerged between full and part estimates, suggesting that the relative parameter values are resistant to the effects of concatenation. A possible exception is seen in the case of joint model 'beta' parameter estimates, which appear to be almost fully dampened by concatenation. This will need to be kept in mind when interpreting the corresponding estimates in the main analyses. Notwithstanding this issue, the canonical correlation analysis revealed strong correlations between parameter estimates for long and short sequences for both simulation sets for the adaptation model (set 2a: Wilks' lambda = 6.56e-06; $F(16,345.86) = 1{,}053.46$; $p < 0.001$; set 2b: Wilks' lambda = 1.08e-05; $F(16,345.86) = 890.62$; $p < 0.001$) and the joint model (set 2a: Wilks' lambda = 1.24e-04; $F(16, 44705.19) = 50{,}307.49$; $p < 0.001$; set 2b: Wilks' lambda = 7.83e-04; $F(16, 44705.19) = 26{,}246.55$; $p < 0.001$). Based on the results of both sets of simulations, we conclude that the reliability of ADAM parameter estimates is not seriously undermined by differing sequence lengths or by concatenation.

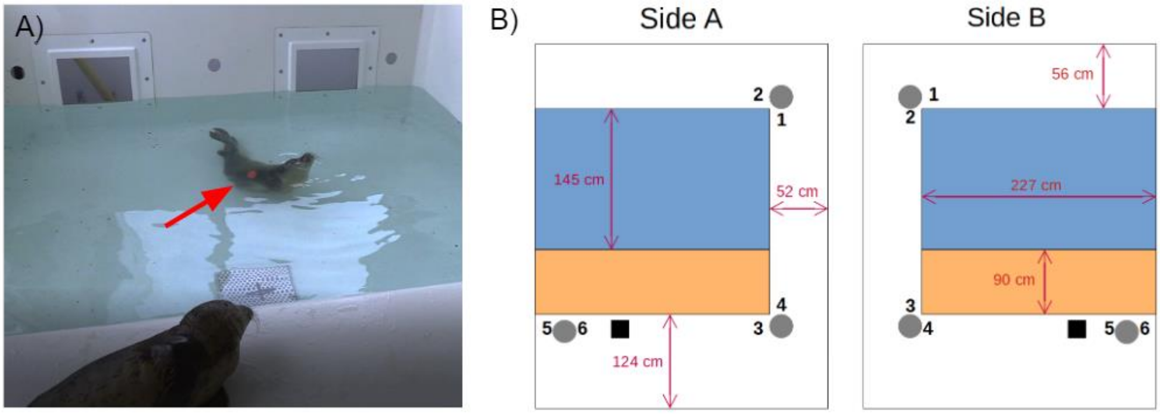

Figure S1. Housing conditions and experimental setup for the 2-way interaction condition. (A) One of the two pups has a red mark (denoted by the red arrow) to facilitate identification during video analyses. (B) Two placement configurations of microphones and tripods in the intensive care units. Grey circles: tripods each holding 2 microphones (which are labelled as 1–6); black square: video camera; blue rectangle: pool with water; orange rectangle: plateau.

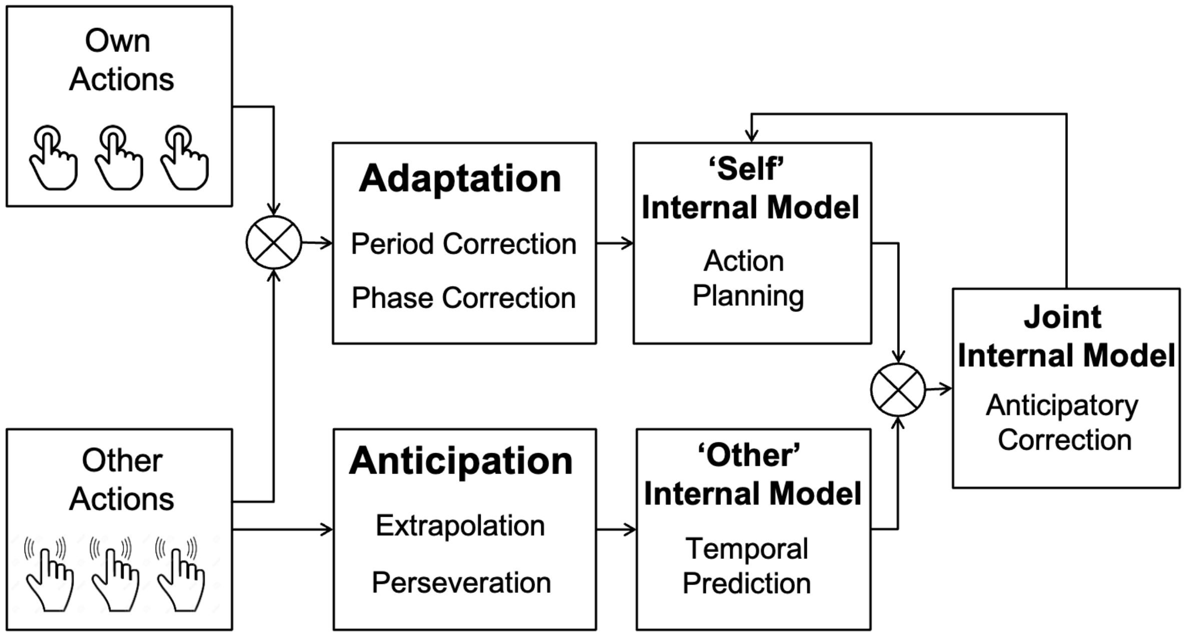

Figure S2. Schematic of the ADaptation and Anticipation Model (ADAM) of sensorimotor coordination applied to interpersonal coordination. The synchronisation of one's own actions with another's actions is facilitated by temporal adaptation mechanisms that influence action planning in the 'self' internal model, anticipation mechanisms that enable temporal prediction in the 'other' internal model, and an anticipatory error correction mechanism that reduces discrepancies between plans and predictions in a joint internal model before a motor command is issued.

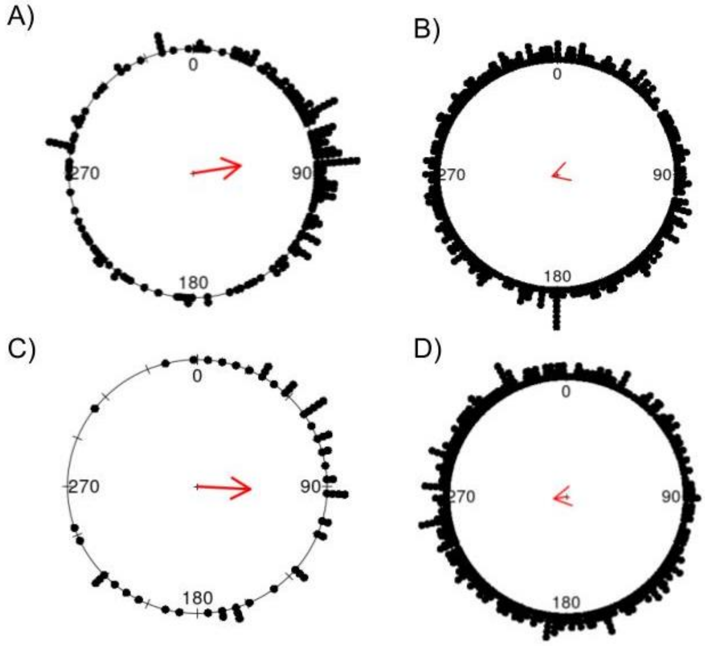

Figure S3. Circular plots for response phases in all four behavioural contexts. (A) 1-way interaction. (B) Alone. (C) 2-way interaction. (D) Silent partner. Angles are measured in degrees starting from 0˚ and going clockwise to 360˚. The data points are shown around the circumference in black, while the red arrows indicate the circular mean (μ). The length of the arrow corresponds to the value of the mean resultant length (ρ).

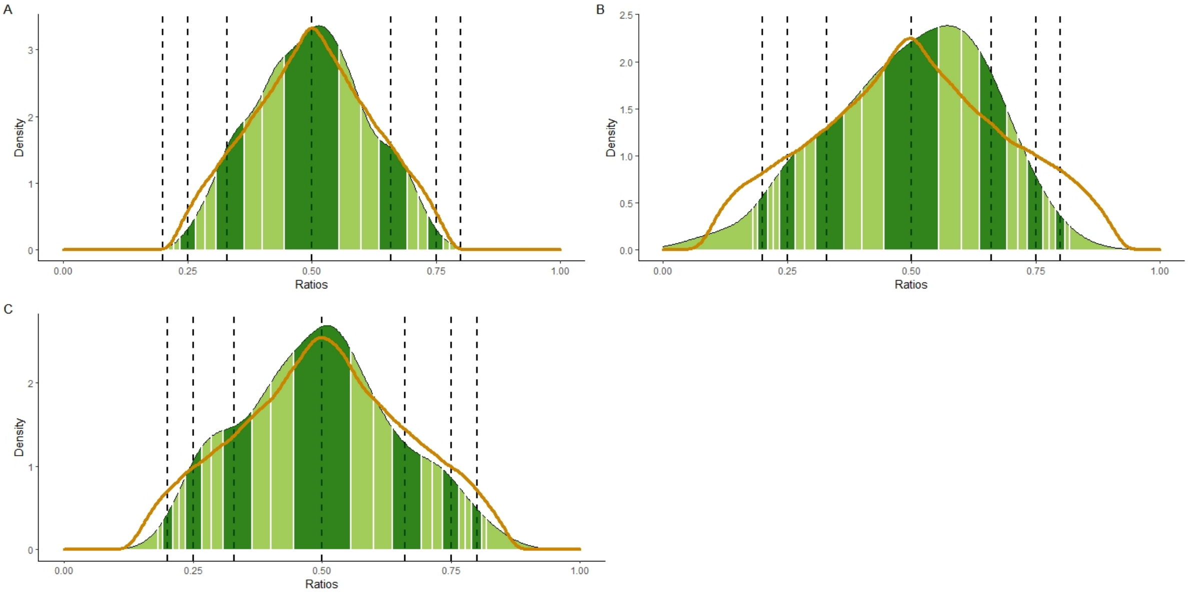

Figure S4. IOI ratio density plots for pups (A) C, (B) E, and (C) H in the silent partner context. See Figure 3 for details. For all three pups, the empirical ratio distribution does not significantly differ from the simulated ratio distribution. Note that the scale of the y-axes differs.

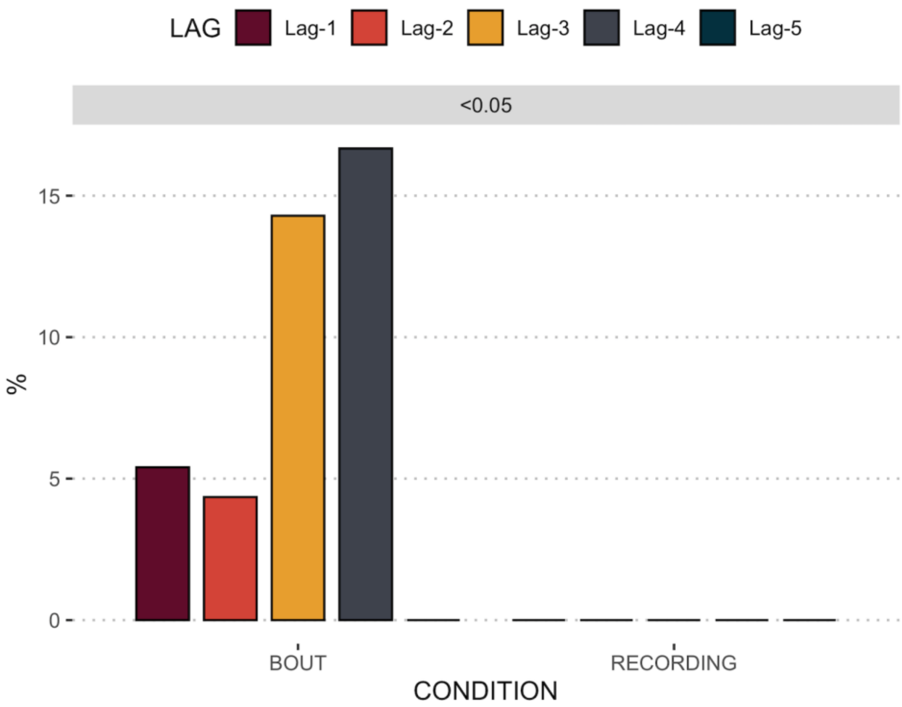

Figure S5. Granger causality results for the 1-way interaction context. The bar plots indicate the percentage of times (y-axis) that the playback timing significantly influenced ($p < 0.05$) the timing of the pup call for the different lag values and the two conditions we considered (different bouts within a recording vs whole recording).

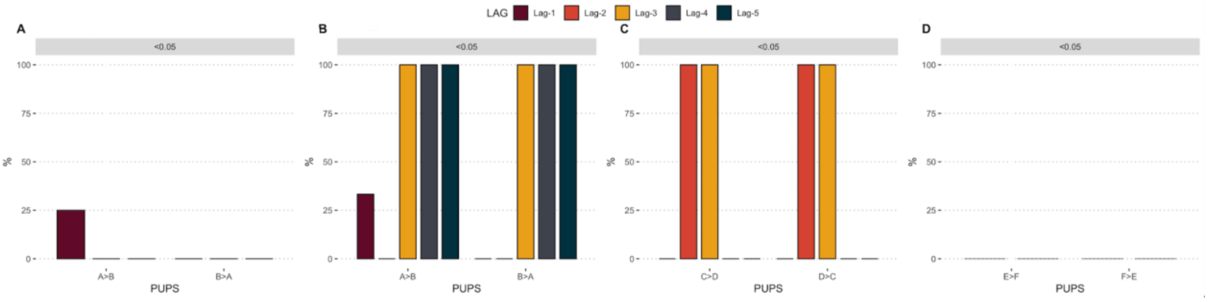

Figure S6. Granger causality results for the 2-way interaction context. The bar plots indicate the percentage of times (y-axis) that the timing of a calling pup was significantly influenced ($p < 0.05$) by another calling pup (i.e., the partner). We considered different pairs of pups bidirectionally and different lag values. We measured Granger Causality for pup pair A/B in two conditions: (A) different bouts within a recording and (B) whole recordings. For pup pairs C/D (C) and E/F (D) we measured Granger Causality only on whole recordings.

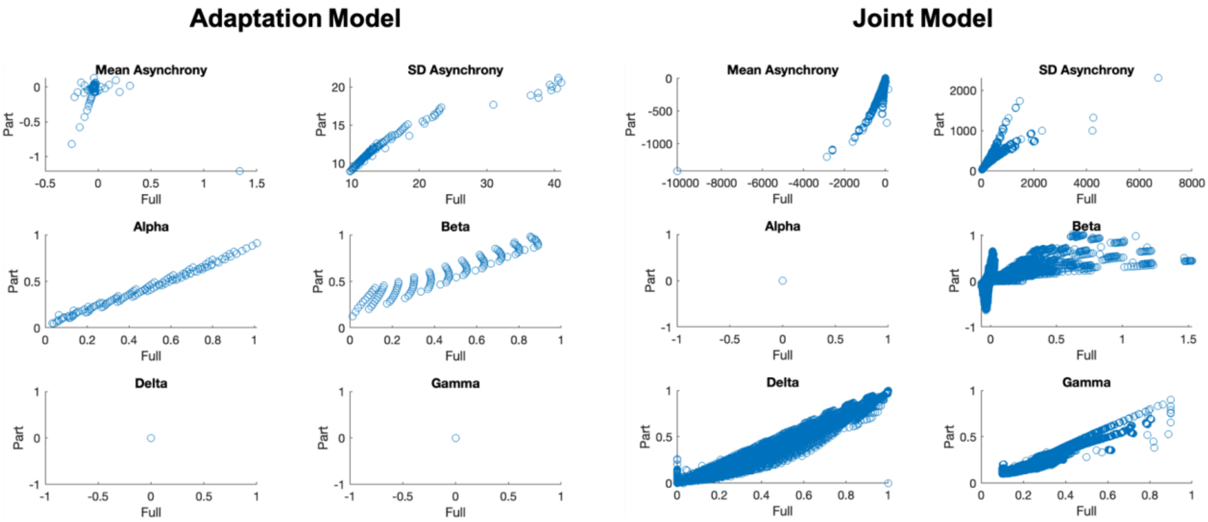

Figure S7. ADAM set 1a results. Scatter plots showing relationships between simulated long (full) and short (part) sensorimotor synchronisation time series for multiple sequences (set 1a) on performance measures (mean asynchrony and the standard deviation of asynchronies) and parameter estimates (alpha, beta, delta, and/or gamma) from the adaptation only model (left) and full 'joint' model (right) versions of ADAM.

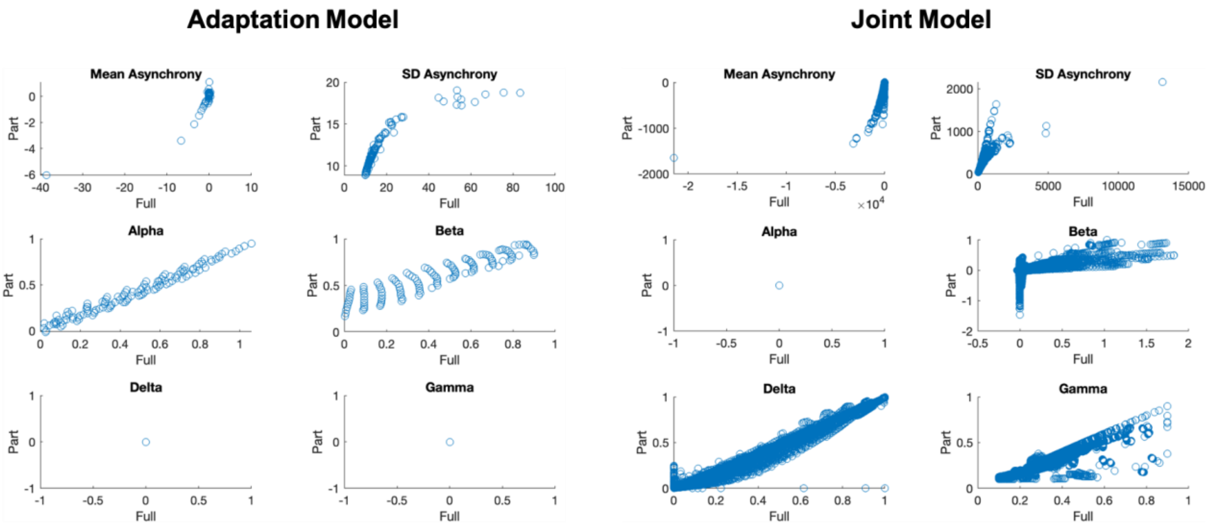

Figure S8. ADAM set 1b results. Scatter plots showing relationships between simulated long (full) and short (part) sensorimotor synchronisation time series for a single sequence (set 1b) on performance measures (mean asynchrony and the standard deviation of asynchronies) and parameter estimates (alpha, beta, delta, and/or gamma) from the adaptation only model (left) and full 'joint' model (right) versions of ADAM.

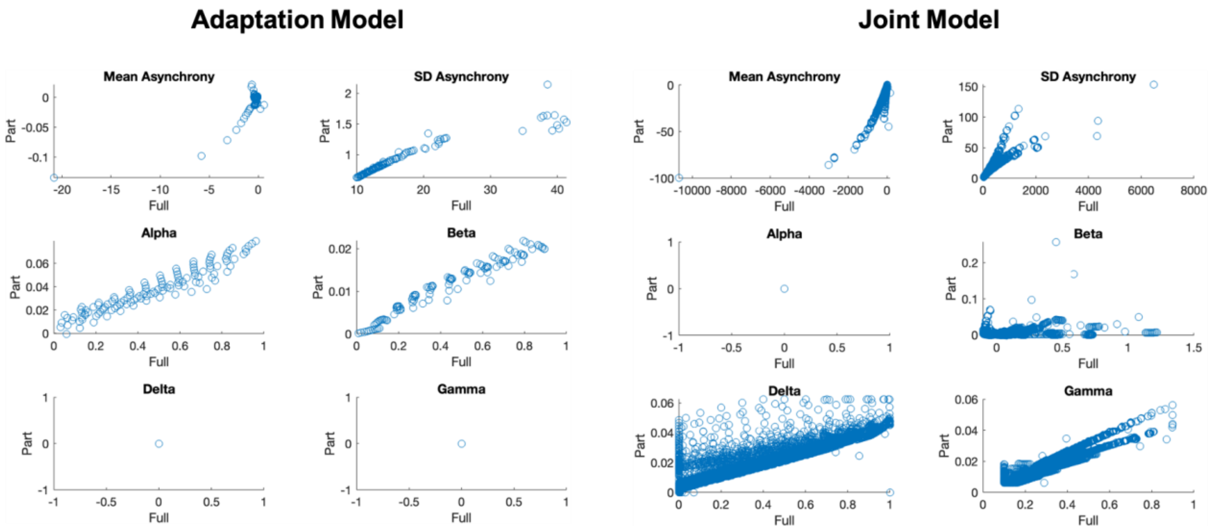

Figure S9. ADAM set 2a results. Scatter plots showing relationships between simulated long (full) and concatenated short (part) sensorimotor synchronisation time series for multiple sequences (set 2a) on performance measures (mean asynchrony and the standard deviation of asynchronies) and parameter estimates (alpha, beta, delta, and/or gamma) from the adaptation only model (left) and full 'joint' model (right) versions of ADAM.

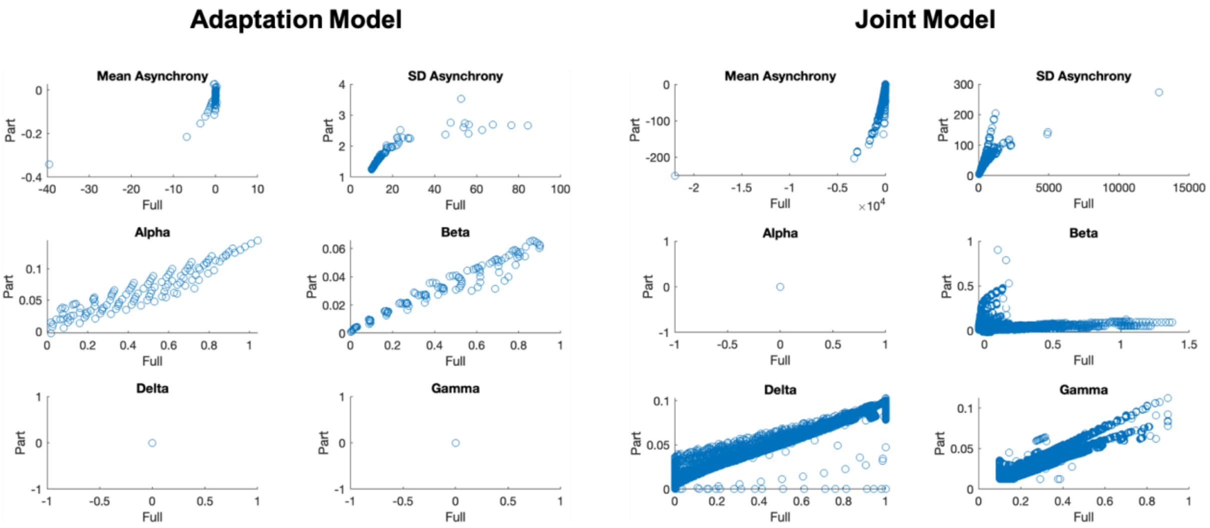

Figure S10. ADAM set 2b results. Scatter plots showing relationships between simulated long (full) and concatenated short (part) sensorimotor synchronisation time series for a single sequence (set 2b) on performance measures (mean asynchrony and the standard deviation of asynchronies) and parameter estimates (alpha, beta, delta, and/or gamma) from the adaptation only model (left) and full 'joint' model (right) versions of ADAM.

Table S1. Dataset overview for four behavioural contexts.

Behavioural context	*Number of pups*	*Pup ID(s)*	*Acoustic stimuli*	*Number of sessions*	*Dataset reference*
Alone	1	I	None	20	[1]
Silent partner	8	A/B C/D E/F G/H	None	3 1 2 2	Unpublished (this study)
1-way interaction	1	I	Playback calls	5	[2]
2-way interaction	8	A/B C/D E/F G/H	Partner calls	3 1 2 2	Unpublished (this study)

Table S2. The contribution of each pup in each analysis is shown with an X.

Pup	*Categorical rhythms analysis*	*Granger causality*	*ADAM*	*Circular statistics*	**TOTAL**
A	X	X	X	X	**4**
B	X	X	X	X	**4**
C	X	X		X	**3**
D		X		X	**2**
E	X	X		X	**3**
F		X		X	**2**
G				X	**1**
H	X			X	**2**
I	X	X	X	X	**4**
TOTAL	**6**	**7**	**3**	**9**	

Table S3. Descriptive statistics of response phases for each behavioural context. n = total number of pups; N = number of recorded calls; μ = mean direction; CI = 95% confidence intervals for μ; v = circular standard deviation; ρ mean resultant length.

Behavioural context	*Pup ID & total pup number (n)*	*N*	*Median*	μ	*CI*	v	ρ
Alone	I (n = 1)	752	253.53°	252.01°	182.60° – 321.50°	2.51°	0.04
Silent partner	A-H (n = 8)	790	258.93°	260.33°	233.40° – 287.70°	2.12°	0.11
1-way interaction	I (n = 1)	205	80.91°	80.98°	68.92° – 93.01°	1.38°	0.38
2-way interaction	A-H (n = 8)	57	91.81°	92.91°	66.48° – 117.90°	1.34°	0.41

Table S4. Descriptive statistics of response phases and Rayleigh test results for each pup were recorded in the silent partner context. *N* = number of calls recorded; μ = mean direction; v = circular standard deviation; *z* = Rayleigh test statistic; *p* = p-value. Significant p-values are bolded.

Pup ID	*N*	μ	v	*z*	*p*
A	65	119.35°	2.31°	0.07	0.727
B	59	227.42°	1.49°	0.33	**0.002**
C	474	280.46°	2.07°	0.12	**0.002**
D	10	325.69°	1.41°	0.37	0.258
E	42	215.85°	1.69°	0.24	0.0887
F	12	256.90°	1.76°	0.21	0.5967
G	2	195.52°	1.28°	0.44	0.7363
H	126	252.57°	2.39°	0.06	0.6604

Table S5. von Mises distribution assumption test results based on a significance level of 0.01. n = total pup number; κ = concentration parameter.

Dataset	*Pup IDs*	*Test score*	*k*	*Outcome*
Alone	I	0.04	0.08	von Mises distributed
Silent partner	A - H	0.07	0.08	von Mises distributed
1-way interaction	I	0.21	0.11	NOT von Mises distributed
2-way interaction	A - H	0.07	0.11	von Mises distributed

Table S6. Additional tests for circular uniformity in the 1-way interaction context, which was not von Mises distributed. Rao's spacing, Kuiper's V, and Watson's U^2 test validated the Rayleigh test results. Significant p-values are bolded.

Test	*Statistic*	*p*
Rao's Spacing	163.0	< **0.01**
Watson's	1.81	< **0.01**
Kuiper's	4.46	< **0.01**

Table S7. On- and off-integer ratio bin boundaries used in the categorical rhythm analysis. Fractions denote the left and right boundaries of each on- and off-ratio bin. For on-integer ratios, the ratio itself is also given in parentheses in the 'Centre' column.

Left bound	*Centre*	*Right bound*
1/5.75	1/5.5 (off)	1/5.25
1/5.25	1/5 (on, 1:4)	1/4.75
1/4.75	1/4.5 (off)	1/4.25
1/4.25	1/4 (on, 1:3)	1/3.75
1/3.75	1/3.5 (off)	1/3.25
1/3.25	1/3 (on, 1:2)	1/2.75
1/2.75	1/2.5 (off)	1/2.25
1/2.25	1/2 (on, 1:1)	1.25/2.25
1.25/2.25	1.5/2.5 (off)	1.75/2.75
1.75/2.75	2/3 (on, 2:1)	2.25/3.25
2.25/3.25	2.5/3.5 (off)	2.75/3.75
2.75/3.75	3/4 (on, 3:1)	3.25/4.25
3.25/4.25	3.5/4.5 (off)	3.75/4.75
3.75/4.75	4/5 (on, 4:1)	4.25/5.25
4.25/5.25	4.5/5.5 (off)	4.75/5.75

Table S8. Range of empirically observed IOIs for each tested pup/condition. Pups D (n=10), F (n=12), and G (n=2) had too few ratios to analyse.

Pup ID	*Number of ratios*	*Behavioural context*	*Corresponding part of Figure 1*	*Observed IOI range (s)*
I	752	Alone	A	0.341–4.121
A	65	Silent partner	B	0.711–4.119
B	59			1.130–4.411
C	474			0.927–3.429
E	42			0.644–7.534
H	126			0.763–5.189
I	145	1-way interaction (disregarding playback)	C (blue bars)	0.034–4.689
I	205	1-way interaction (pup I responds)	C (grey bars)	
A	16	2-way interaction (pup A responds)	D (grey bars, purple shaded interactions)	0.144–4.355
B	22	2-way interaction (pup B responds)	D (grey bars, yellow shaded interactions)	0.150–2.245

Table S9. Categorical rhythm analysis one-sample Kolmogorov-Smirnov (KS) and paired Wilcoxon signed-rank (WSR) test results for each pup/behavioural context. The KS tests assessed whether empirical and simulated ratio distributions significantly differed (*D* = D statistic range; *p* = p-value), and the WSR tests assessed whether empirical ratio distributions exhibited significant peaks at seven tested ratios (Med = pseudo-median; dashes = ratio sample sizes too small to perform test, i.e., 95% confidence interval could not be estimated). The 1:4 and 1:3 ratio WSR test results are omitted because ratio sample sizes were too small for all tested seals. Only conditions with a significant KS test result underwent categorical rhythm analyses. Significant p-values are coded as: 0 = ***, 0.001 = **, and 0.01 = *.

Pup ID	*Behavioural context*	*KS test*		*WSR test*									
		D	*Mean p*	*1:2*		*1:1*		*2:1*		*3:1*		*4:1*	
				Med	*p*	*Med*	*p*	*Med*	*p*	*Med*	*p*	*Med*	*p*
I	Alone	0.117–0.197	***	3.25	0.611	-6.75	0.091	-2.75	0.215	-3.75	0.103	-	-
A	Silent partner	0.077–0.385	0.304										
B		0.068–0.339	0.473										
C		0.025–0.114	0.495										
E		0.071–0.381	0.584										
H		0.048–0.230	0.535										
I	1-way interaction (disregarding playback)	0.048–0.166	0.609										

I	1-way interaction (pup I responds)	0.376–0.590	***	-	-	2.25	0.505	11.0	0.245	-21.3	0.068	5.25	0.712
A	2-way interaction (A responds)	0.125–0.688	0.440										
B	2-way interaction (B responds)	0.273–0.773	*	-	-	5.73	0.414	-36.0	0.099	-	-	-	-

Table S10. Categorical rhythm analysis KS test results for each pup, across conditions. All *p*-values were significant (i.e., Bonferonni corrected to $p \leq 0.017$ for pup I analyses; $p \leq 0.05$ for pup A and B analyses) and are coded as: 0 = ***, 0.001 = **, and 0.01 = *.

Pup ID	*Behavioural context*	*D statistic*	*p*
I	Alone vs. 1-way interaction (disregarding playback)	0.192	***
	Alone vs. 1-way interaction (pup I responds)	0.605	***
	1-way interaction (disregarding playback) vs. 1-way interaction (pup I responds)	0.464	***
A	Silent partner vs. two-way interaction (pup A responds)	0.375	0.04
B	Silent partner vs. two-way interaction (pup B responds)	0.705	***

Table S11. Granger causality test results for both behavioural contexts (1-way and 2-way). F- and p-values are provided. We reported bouts and sessions for which at least one of the tested lags yielded statistically significant results (in bold): four bouts featuring pup I; one bout and two sessions featuring pups A and B; one session featuring pups C and D. Dashes denote cases where the test was inconclusive due to small sample size [14].

			Lag-1		*Lag-2*		*Lag-3*		*Lag-4*		*Lag-5*	
Pup ID	*Behavioural context*	*Bout*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
I	1-way interaction	7 (n=16)	3.132	0.102	6.934	**0.015**	5.634	**0.035**	2.869	0.206	-	-
		10 (n=7)	89.601	**0.002**	-	-	-	-	-	-	-	-
		13 (n=14)	0.017	0.900	0.692	0.532	0.486	0.710	4818.6	**0.011**	-	-
		30 (n=10)	8.326	**0.028**	4.907	0.113	-	-	-	-	-	-
A	2-way interaction (B responds)	3 (n=11)	10.601	**0.014**	3.626	0.126	0.883	0.635	-	-	-	-
Pup ID	*Behavioural context*	*Session*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
A	2-way interaction (B responds)	1 (n=71)	0.709	0.403	0.367	0.694	5.276	**0.003**	3.179	**0.020**	3.255	**0.012**
B	2-way interaction (A responds)		0.111	0.740	0.028	0.972	4.786	**0.005**	2.562	**0.047**	2.584	**0.036**
A	2-way interaction (B responds)	2 (n=8)	8.369	**0.044**	1.922	0.454	-	-	-	-	-	-
C	2-way interaction (D responds)	4 (n=17)	0.160	0.695	16.427	**0.001**	8.461	**0.010**	0.533	0.721	1.548	0.542
D	2-way interaction (C responds)		0.164	0.692	16.413	**0.001**	8.444	**0.010**	0.531	0.723	1.870	0.503

Table S12. ADAM descriptive statistics for asynchronies, IOIs, and the number of events for sequences for all pairs of time series analysed. Data are included for the 2-way interaction context between pups A and B, and for pup I in five playback sessions (PB1-PB5). Median, Min, and Max asynchrony are computed with unsigned (absolute) asynchronies; SD of asynchronies is computed with signed asynchronies.

Pup pair	*Asynchronies (s)*				*Inter-onset intervals (s)*				*N events*
	Median	*SD*	*Min*	*Max*	*Median*	*SD*	*Min*	*Max*	
A/B	0.454	3.243	0.000	21.003	7.689	14.610	1.215	50.054	47
I/PB1	0.731	0.696	0.156	2.602	8.265	12.482	1.847	47.964	22
I/PB2	0.624	1.712	0.005	10.226	6.064	10.235	1.105	55.608	60
I/PB3	0.553	0.916	0.105	4.028	16.683	15.248	0.679	53.986	21
I/PB4	0.617	1.166	0.041	8.399	4.510	7.995	0.504	34.246	91
I/PB5	0.575	1.753	0.081	11.842	4.564	8.925	0.724	51.932	86

Table S13. Significance test results for parameter estimates from the Adaptation Model version of ADAM for seal pups in the one-way playback sessions (1-5) and two-way interaction behavioural context condition (with bi-directional focal and reference 'ref' roles). Parameter estimates were converted to z scores based on permuted data estimates (Figure 4), and then p-values were computed for the z scores. 2-tailed $p < 0.05$ = **, 1-tailed $p < 0.10$ = *.

Pup ID	*ADAM parameter*	*Behavioural context*	*Z score*	*p*
292	Phase correction	1-way interaction: Playback 1	0.682	0.495
		1-way interaction: Playback 2	0.359	0.720
		1-way interaction: Playback 3	-0.271	0.786
		1-way interaction: Playback 4	-2.065	0.039*
		1-way interaction: Playback 5	0.840	0.401
	Period correction	1-way interaction: Playback 1	-1.121	0.262
		1-way interaction: Playback 2	-0.760	0.447
		1-way interaction: Playback 3	-0.504	0.614
		1-way interaction: Playback 4	-0.948	0.343
		1-way interaction: Playback 5	-1.529	0.126
108	Phase correction	2-way interaction (108 focal, 112 ref)	-0.493	0.622
	Period correction	2-way interaction (108 focal, 112 ref)	2.758	0.006**
112	Phase correction	2-way interaction (112 focal, 108 ref)	1.687	0.092*
	Period correction	2-way interaction (112 focal, 108 ref)	-2.545	0.011**

Table S14. Significance test results for parameter estimates from the Joint Model version of ADAM for seal pups in the playback sessions (1-5) and two-way interaction behavioural context condition (with bi-directional focal and reference ‘ref’ roles). Parameter estimates were converted to z scores based on permuted data estimates (Figure 4), and then p-values were computed for the z scores. 2-tailed $p < 0.05$ = **, 1-tailed $p < 0.10$ = *.

Pup ID	*ADAM parameter*	*Behavioural context*	*Z score*	*p*
292	Period correction	1-way interaction: Playback 1	-1.977	0.048**
		1-way interaction: Playback 2	-0.618	0.537
		1-way interaction: Playback 3	-1.582	0.114
		1-way interaction: Playback 4	-1.019	0.308
		1-way interaction: Playback 5	-0.916	0.360
	Prediction-tracking	1-way interaction: Playback 1	-2.263	0.024**
		1-way interaction: Playback 2	-2.158	0.031**
		1-way interaction: Playback 3	0.897	0.370
		1-way interaction: Playback 4	-1.909	0.056*
		1-way interaction: Playback 5	-1.006	0.314
108	Period correction	2-way interaction (108 focal, 112 ref)	5.399	< 0.001**
	Prediction-tracking	2-way interaction (108 focal, 112 ref)	1.348	0.178
112	Period correction	2-way interaction (112 focal, 108 ref)	-2.622	0.009**
	Prediction-tracking	2-way interaction (112 focal, 108 ref)	-0.571	0.568

Supplemental References

1. Ravignani A, Kello CT, de Reus K, Kotz SA, Dalla Bella S, Méndez-Aróstegui M, Rapado-Tamarit B, Rubio-Garcia A, de Boer B. 2019 Ontogeny of vocal rhythms in harbour seal pups: an exploratory study. *Curr Zool.* 65(1), 107–120. (doi: 10.1093/cz/zoy055)
2. Ravignani A. 2019 Timing of antisynchronous calling: A case study in a harbour seal pup (Phoca vitulina). *J Comp Psychol.* 133(2), 272–277. (doi:10.1037/com0000160)
3. Boersma P, Weenink D. 2021 Praat: doing phonetics by computer (Computer program). Version 6.1.24. (http://www.praat.org/)
4. Buschmeier H, Wlodarczak M. 2013 TextGridTools: A TextGrid Processing and Analysis Toolkit for Python. *Tagungsband der 24. Konferenz zur Elektronischen Sprachsignalverarbeitung (ESSV 2013)*. Bielefeld, Germany, 152–157.
5. Mazzoni D, Dannenberg R. 1999 Audacity: a free, cross-platform digital audio editor. (Computer program). Version 2.3.3. (http://audacityteam.org/)
6. Friard O, Gamba M. 2016 BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*. 7(11), 1325–1330. (doi:10.1111/2041-210X.12584)
7. R Core Team. 2019 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (https://www.R-project.org/)
8. RStudio Team. 2020 RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. (http://www.rstudio.com/)
9. Lund U, Agostinelli C, Agostinelli MC. 2017 circular: Circular Statistics. *Repository CRAN*, 1–142.
10. Cremers J, Klugkist I. 2018 One direction? A tutorial for circular data analysis using R with examples in cognitive psychology. *Front. Psychol.* 9, 2040. (doi: 10.3389/fpsyg.2018.02040)
11. Pewsey A, Neuhäuser M, Ruxton GD. 2013 *Circular statistics in R*. Oxford University Press.
12. Landler L, Ruxton GD, Malkemper EP. 2018 Circular data in biology: advice for effectively implementing statistical procedures. *Behav. Ecol. Sociobiol.* 72(8), 1–10. (doi:10.1007/s00265-018-2538-y)
13. Granger CWJ. 1969 Investigating causal relations by econometric models and cross-spectral methods. *Econometrica.* 37, 424–438. (doi: 10.2307/1912791)
14. Zeileis A, Hothorn T. 2002 Diagnostic Checking in Regression Relationships. *R News*, 2(3), 7–10. (https://CRAN.R-project.org/doc/Rnews/)
15. Trapletti A, Hornik K. 2022. *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-52. (https://CRAN.R-project.org/package=tseries)
16. Kassambara A. 2020 ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. (https://CRAN.R-project.org/package=ggpubr)
17. van der Steen MC, Jacoby N, Fairhurst MT, Keller PE. 2015 Sensorimotor synchronization with tempo-changing auditory sequences: Modeling temporal adaptation and anticipation. *Brain Res*. 1626, 66–87. (doi:10.1016/j.brainres.2015.01.053)
18. Jacoby N, Repp BH. 2012 A general linear framework for the comparison and evaluation of models of sensorimotor synchronization. *Biol. Cybern.* 106(3), 135-154. (doi:10.1007/s00422-012-0482-x)
19. Repp BH, Keller PE, Jacoby N. 2012 Quantifying phase correction in sensorimotor synchronization: empirical comparison of three paradigms. *Acta Psychol (Amst)*. 139(2), 281-290. (doi:10.1016/j.actpsy.2011.11.002)
20. Jacoby N, Tishby N, Repp BH, Ahissar M, Keller PE. 2015 Parameter estimation of linear sensorimotor synchronization models: Phase correction, period correction, and ensemble synchronization. *Timing & Time Perception*. 3(1-2), 52–87. (doi:10.1163/22134468-00002048)
21. Jacoby N, Keller PE, Repp BH, Ahissar M, Tishby N. 2015 Lower bound on the accuracy of parameter estimation methods for linear sensorimotor synchronization models. *Timing & Time Perception*. 3(1-2), 32-51. (doi:10.1163/22134468-00002047)